Graphical Rule-Based Representation of Signal-Transduction Networks

James R. Faeder
Theoretical Biology & Biophysics
Theoretical Division
Los Alamos Natl Lab
faeder@lanl.gov

Michael L. Blinov
Theoretical Biology & Biophysics
Theoretical Division
Los Alamos Natl Lab
mblinov@lanl.gov

William S. Hlavacek*
Theoretical Biology & Biophysics
Theoretical Division
Los Alamos Natl Lab
wish@lanl.gov

ABSTRACT

The process by which a cell senses and responds to its environment, as in signal transduction, is often mediated by a network of protein-protein interactions, in which proteins combine to form complexes and undergo post-translational modifications, which regulate their enzymatic and binding activities. A typical signaling protein contains multiple sites of protein interaction and modification and may contain catalytic domains. As a result, interactions of signaling proteins have the potential to generate a combinatorially large number of complexes and modified states, and representing signal-transduction networks can be challenging. Representation, in the form of a diagram or model, usually involves a tradeoff between comprehensibility and precision: comprehensible representations tend to be ambiguous or incomplete, whereas precise representations, such as a long list of chemical species and reactions in a network, tend to be incomprehensible. Here, we develop conventions for representing signal-transduction networks that are both comprehensible and precise. Labeled nodes represent components of proteins and their states, and edges represent bonds between components. Binding and enzymatic reactions are described by reaction rules, in which left graphs define the properties of reactants and right graphs define the products that result from transformations of reactants. The reaction rules can be evaluated to derive a mathematical model.

Categories and Subject Descriptors

I.6.5 [**Simulation and Modeling**]: Model Development—*modeling methodologies*; E.1 [**Data Structures**]: Graphs and networks; J.3 [**Life and Medical Sciences**]: Biology and genetics

Keywords

Signal Transduction, Biological Systems Modeling, Protein Interactions, Graph Isomorphism, Graph Rewriting

*Author to whom correspondence should be addressed: Mail Stop K710, T-10, Los Alamos National Laboratory, Los Alamos, NM 87545, USA. Tel: 505-665-1355 Fax: 505-665-3493

SAC '05, March 13–17, 2005, Santa Fe, New Mexico, USA.

1. INTRODUCTION

Many cellular responses to environmental signals are mediated by networks of interacting proteins that detect signals (e.g., ligands of cell-surface receptors) and transduce these signals into responses, such as the release of stored factors, changes in gene expression, and cell movement, proliferation, differentiation, or death. After the introduction of a signal, the proteins in a signal-transduction network typically undergo post-translational modifications (e.g., tyrosine phosphorylation), which affect their binding and enzymatic activities, and concurrently combine to form a variety of heterogeneous complexes [1, 2]. These complexes, which are often transient and prominent in the vicinity of the inner cell membrane, regulate enzymatic activities, for example, by serving to co-localize enzymes and substrates, which is a common mechanism for controlling enzyme specificity [3]. The number of protein complexes and modification states that potentially can be generated during the response to a signal is combinatorially large and generally far greater than the number of proteins involved in signal transduction, because signaling proteins contain multiple sites of modification and may interact with multiple binding partners [4, 5].

There are at least two reasons to account for all the possible protein states and complexes in a signal-transduction network, as numerous as these may be. First, most states and complexes may be unimportant, but in general, it is impossible to determine intuitively which are the important ones from knowledge of pairwise protein interactions, which is the usual level of detail available, even for a well-studied system. Second, the catalytic activities of signaling proteins are highly regulated by molecular context. For example, the activity of a protein tyrosine kinase (PTK) might depend on the phosphorylation state of its activation loop and its specificity might depend on the proximity of a specific substrate. Thus, we desire representations of signal-transduction networks that precisely account for the full array of possible protein states and complexes implied by a given set of protein interactions. To make practical use of these representa-

tions, one must be able to translate them into mathematical and computational models, which can then be used to interpret data, predict the behavior of a system, and design experiments to test model-based predictions [6, 7, 8].

A precise representation of a signal-transduction network can be provided by a comprehensive list of the chemical species and reactions in the network [9]. However, this type of representation is difficult to comprehend, even for small systems, in that it obscures the underlying protein interactions that give rise to the chemical species and reactions. The list may also be quite long. A more comprehensible type of representation, and one that is commonly used, is provided by a diagrammatic interaction map in which proteins and their interactions (or the functional consequences of these interactions) are indicated by labeled cartoons and arrows. Formal conventions have been proposed for drawing interaction maps such that they have precise meanings [10, 11, 12, 13]. However, interaction maps tend to suffer from a tradeoff between precision and comprehensibility. Maps that are precise enough to have an unambiguous mathematical interpretation may be no more understandable than a list of reactions. On the other hand, ad hoc heuristic maps, which are more the norm, may clearly illustrate certain aspects of a system but are ambiguous and lack a mathematical interpretation.

One way to achieve a precise and understandable representation involves the specification of a reaction rule for each type of protein-protein interaction in a network [5, 8, 14, 15, 16]. In this approach, strings are used to represent chemical species and regular expressions are used to represent groups of chemical species with particular attributes. Reaction rules, or generalized reactions, are written in the same form as a chemical reaction but regular expressions are allowed. These string-matching patterns identify groups of chemical species by indicating the shared attributes of a group. Thus, the rules can be used to find, through string matching, the chemical species among a set of species that qualify as reactants. The rules also define transformations of reactants into products by providing a rate law and indicating how strings representing reactants should be modified to obtain products. Thus, they are generators of reactions and products, which may include new species. The result of rule application is a list of chemical species and reactions implied by the rules and the seed set of species to which the rules are initially applied. This approach has been used to model early events in signaling by FcεRI [14, 15], a prototypical antigen recognition receptor of the immune system, and to derive preliminary models for an array of other systems [16]. The number of rules that must be specified is comparable to the number of components of proteins in the network, which is usually much less than the number of chemical species.

Here, we extend the rule-based approach described above by defining conventions for using graphs to represent chemical species and groups of chemical species. The introduction of graphs is a natural generalization of the string representation of Blinov et al. [16]. With it, we gain the ability to explicitly and systematically represent the connectivity of protein components in a complex at the expense of finding graph isomorphisms, instead of simply matching strings, when applying reaction rules. Below, we introduce the conventions of representation, present examples, compare graphical rule-based representation with formal diagrammatic representation, and briefly mention the classical problems of graph isomorphisms that must be solved to translate a set of rules into a model. The method of Blinov et al. [16] and ideas presented here will be elaborated in another publication [17].

2. METHOD OF GRAPHICAL REPRESENTATION

Figure 1 introduces a method of using graphs and graph rewriting rules, or graphical reaction rules, to represent signal-transduction networks. We focus on signal transduction and protein-protein interactions, but the conventions of Fig. 1 can be used to represent other types of cellular systems and biomolecular interactions, such as genetic regulatory networks and protein-DNA or protein-lipid interactions. The method is also illustrated with examples specific to the model of Faeder et al. [15] (Figs. 2–5), a model for bivalent ligand interaction with a bivalent cell-surface receptor (Fig. 6), and a model considered in the review of Aladjem et al. [13] (Fig. 7).

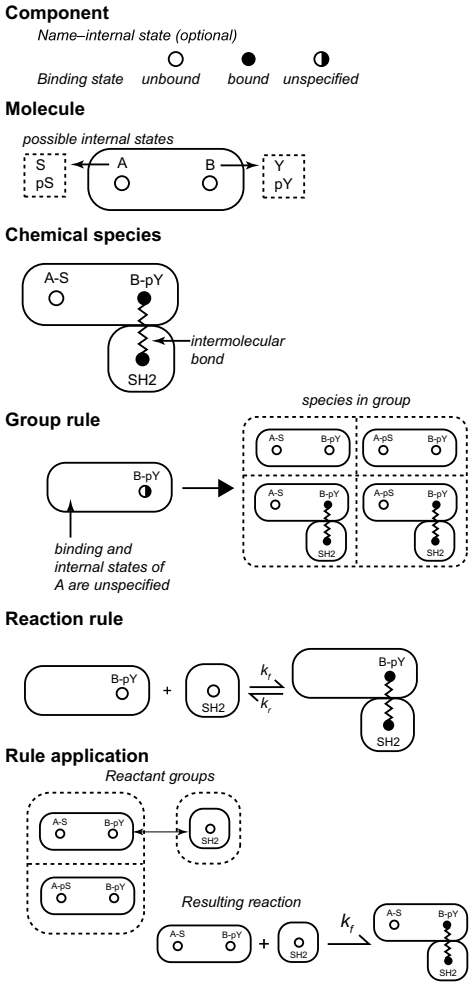

Figure 1: Conventions of the graphical representation.

The method of representation was developed with the following features of signaling proteins in mind. These proteins are generally comprised of conserved modular domains.

Some domains are catalytic. A PTK domain, for example, catalyzes the addition of a phosphate group to a tyrosine residue of a protein substrate. Some domains are responsible for biomolecular recognition; protein interaction domains recognize specific types of sites in proteins and other biomolecules [2]. For example, the Src homology 2 (SH2) domain recognizes phosphorylated tyrosines in protein motifs, such as the immunoreceptor tyrosine-based activation motif (ITAM). The activities of protein domains can be regulated by post-translational modifications, which are catalyzed by signaling proteins. For example, the activity of a PTK domain can be upregulated by autophosphorylation of its activation loop, and the affinity of an SH2 domain for an ITAM can be upregulated by PTK-mediated ITAM phosphorylation. These modifications can be reversed (e.g., a tyrosine can be dephosphorylated by a protein tyrosine phosphatase). Binding events and conformational changes can also affect the activities of signaling proteins. The challenge is to account for interactions among molecules, each potentially having multiple components, each potentially having a binding or catalytic activity that depends on its bound, conformational, or modification state, which can vary.

2.1 Components, Internal States, and Bonds

The elements of a graph are nodes, labels associated with the nodes, and undirected and unlabeled edges that connect nodes (Fig. 1). Nodes represent components (e.g., sites and domains of proteins), which may have multiple internal states (e.g., phosphorylated or unphosphorylated), labels give the names of components and their internal states, and edges represent bonds between components. Here, we limit discussion to edges that are subject to addition or removal in a graph rewriting step, i.e., bonds affected by signaling. Bonds connecting components that are unaffected by signaling are not represented explicitly. Internal states are introduced as needed or desired to represent bound, conformational, or modification states of a component that are not represented otherwise. As illustrated in Figs. 1, 2 and 7, when a component is defined, it is assigned a name and a list of its allowed internal states (if any) is given.

As discussed later, we will sometimes need to specify the connectivity of a node, for example, to write a reaction rule in which a particular component of a reactant must be unbound. Here, we uniformly use an open (filled) circle for a node that is unconnected (connected) to an edge. A half-filled circle is used for a node that may be either connected or unconnected to an edge. Other ways of specifying connectivity are possible. For example, a special node might be introduced to represent an empty space and connected to nodes of components that are unbound.

2.2 Molecules

A molecule is defined as a set of components that can be treated as a unit (Figs. 1, 2, 6 and 7), such as the components of a polypeptide chain or multimeric protein. A molecule is represented graphically by a box surrounding a set of nodes that represents each component of the molecule. Like components, molecules are assigned names, but here, we usually suppress these names to avoid clutter, because molecules can be distinguished by the shapes of their boxes or the names of their components. Names of components are also suppressed in some cases. Names can be suppressed because we adopt the convention that the components of a molecule are represented at fixed relative positions within a box. These conventions for illustrating a model do not affect the underlying graph representation of components, bonds, and molecules. The internal states of a molecule and its connectivity to external components is determined by the attributes of the nodes representing its components. In our examples, every component is part of a molecule.

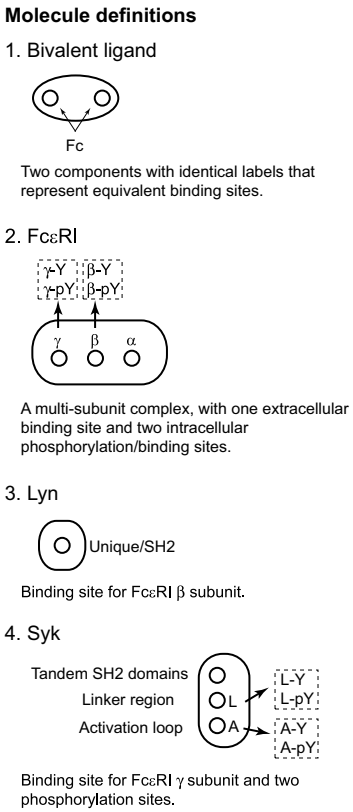

Figure 2: Definition of molecules in the FcεRI model.

2.3 Chemical Species

A chemical species is either a single molecule having all of its components fully defined or a set of connected molecules (i.e., a complex), with each molecule in the set having all of its components fully defined. A component is fully defined if its internal state is specified and its connection with other components is specified. If a component is bound to another component, then the nodes representing the two components are joined by an edge. An example of a chemical species is illustrated in Fig. 1; others are illustrated in Fig. 3. In general, a chemical species is represented by a graph in which nodes are partitioned into molecules, edges connect the nodes of components that are bound to each other, and node labels indicate the particular internal states of those nodes that have multiple allowed states. There is a chemical species for each unique combination of the possible component connections and states in a system.

2.4 Groups of Chemical Species

Groups of chemical species with specified shared features can be defined by graphs that do not completely specify component interactions and states, which we call group rules or

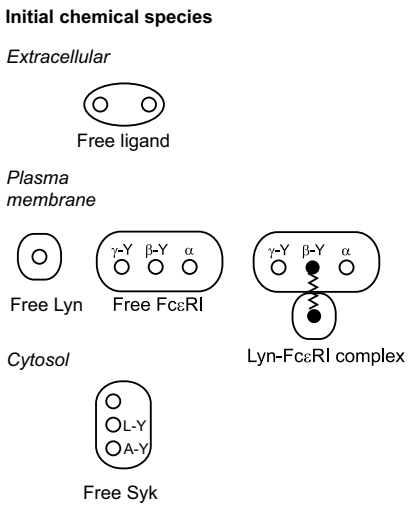

Figure 3: A set of particular chemical species in the FcεRI model.

group graphs. The interactions and states that are specified define the distinguishing features of a group. An example of a group rule is shown in Fig. 1 along with a set of chemical species having features consistent with the rule. In general, given a set of chemical species, this group rule selects all chemical species among the set in which component B of the indicated molecule is in state pY. Because the internal state of component A and the connectivity of B are unspecified in the rule, chemical species selected by the rule can have different states of A and different bound states of B as shown. Additional species, depending on the set of species being tested, could belong to the group, as would be the case if component A was attached to a binding partner in one of the species among the set tested. A second example of a group graph is shown in Fig. 4.

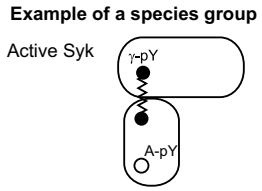

Figure 4: A group of chemical species.

Formally, a chemical species represented by graph X is a member of a group represented by group graph G, if and only if there is a subgraph of X that is isomorphic to G when the internal states are removed from the labels of nodes in X that have unspecified internal states in G. Thus, the problem of identifying which species belong to a group is reduced to the problem of determining whether X contains subgraphs isomorphic to the group graph G, which is known as the subgraph isomorphism problem. A well-known algorithm for finding isomorphic subgraphs is the method of Ullmann [18].

2.5 Reaction Rules

Reaction rules are graph rewriting rules that can be used to generate chemical reactions from a list of chemical species by identifying sets of reactants and defining how reactants are transformed into products. Each rule is comprised of two sets of group graphs (a set of graphs representing reactants and a set of graphs representing products), an arrow pointing from reactants to products, and a rate law. The rate law in general can be any function of the properties of reactants and products, e.g., $k[A][B]$, where k is a rate constant and $[A]$ and $[B]$ are the concentrations of reactants in a biomolecular reaction. An example of a reaction rule is shown in Fig. 1. The bidirectional arrows indicate that the rule is to be applied in both the forward and reverse directions.

The first step in applying a reaction rule to a set of chemical species is to identify the group of species corresponding to each reactant group graph, as described in the previous section. Next, for each combination of reactant species drawn from these groups, the rule is applied by replacing the subgraphs of the reactant species matching the group graphs of reactants with the corresponding group graphs of products to define the products. In carrying out this replacement, component states that are not specified in the product group graphs are not changed. This process of replacing subgraphs of reactants with product group graphs is a graph rewriting step [19], i.e., a cut-and-paste operation (or in some cases, equivalently, a relabeling operation) that transforms reactant graphs in product graphs. The product species that result from graph rewriting are then checked against the current list of chemical species and added to the list if they are not already present. To facilitate this comparison, graphs must be assigned a unique label that does not depend on the order of components, graph partitions (i.e., molecules), or edges. Such labels can be assigned using the canonical graph labeling scheme of McKay [20]. Canonical labels are also useful for checking the generated reaction against the list of previously generated reactions to identify overlaps in reaction rules or to prevent duplication of reactions that are related because of symmetry. An example of application of a reaction rule that would generate two reactions is shown in Fig. 1. The set of rules that generate the model of Faeder et al. [15] is shown in Fig. 5. Other sets of rules are shown in Figs. 6 and 7.

2.6 Generation of a Chemical Reaction Network

An initial set of chemical species must be specified as a starting point for the application of reaction rules and the generation of a chemical reaction network. A typical starting point for network generation would be the set of individual molecules with each component in its resting internal state. A seed set of initial chemical species is shown in Fig. 3; iterative application of the rules of Fig. 5 to this set of species generates the reaction network of Faeder et al. [15], which contains 354 chemical species and 3680 reactions. Iterative application of reaction rules can be carried out until a termination condition is satisfied or all possible species and reactions are generated. An exhaustive generation of all species and reactions accessible from the initial set is a possible termination condition as long as the reaction rules give rise to a finite number of species, but may not be desirable in the case of very large networks, e.g., if the number of chemical species containing a particular molecule exceeds the number of that kind of molecule in a cell. The rules of Fig. 6 provide an example of a rule set for which exhaustive

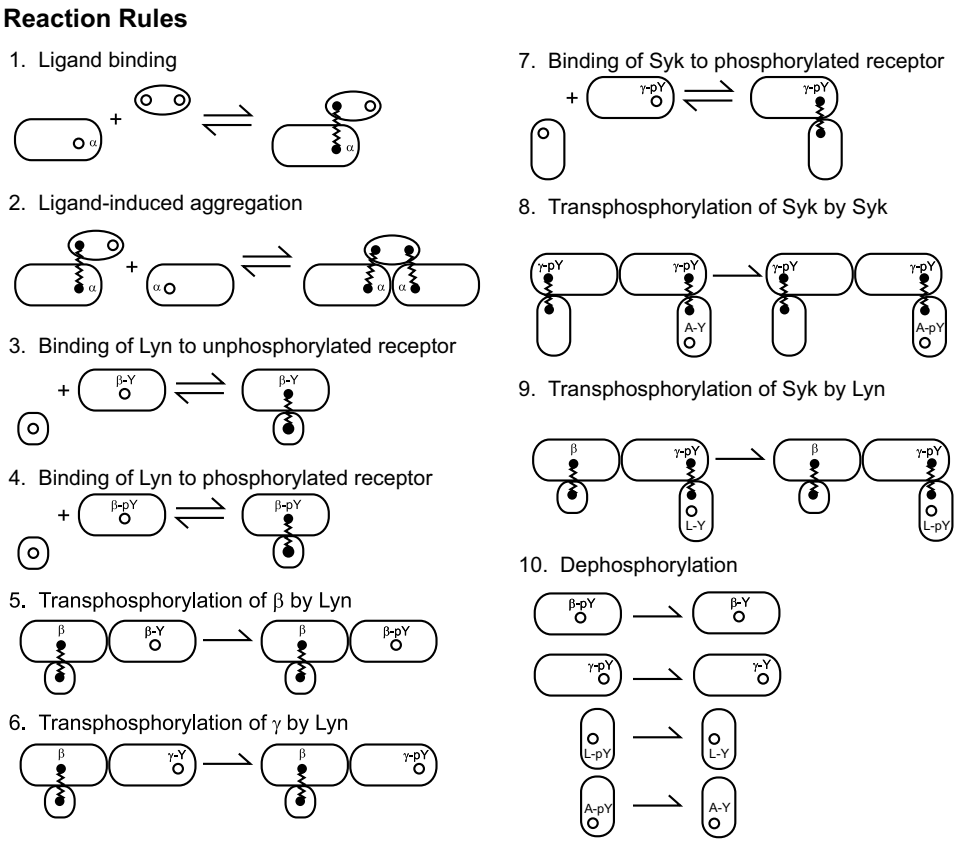

Figure 5: Graphical rule-based representation of the FcεRI model.

generation of all possible chemical species and reactions is impossible. In these cases, other termination conditions are needed. Alternatively, network generation and simulation can proceed in tandem, such that species and reactions are generated on-the-fly as needed [5, 21].

2.7 Output functions

It is often useful to associate a mathematical function with a group of chemical species, such as the sum of concentrations of all members of a group, because experimental observables often correspond to properties of ensembles of chemical species. A group graph and associated function can be specified to calculate this sum. For example, the group rule in Fig. 1 could be used to calculate the concentration of the specified protein phosphorylated on its B domain tyrosine, and the group rule in Fig. 4 could be used to calculate the concentration of receptor-bound autophosphorylated Syk.

3. EXAMPLES

3.1 The FcεRI Network

Figures 2–5 illustrate how the FcεRI signaling model of Faeder et al. [15] can be representated using the proposed graphical conventions. This model vividly illustrates combinatorial complexity. The four molecules of Fig. 2 combined with the ten reaction rules of Fig. 5 imply 354 chemical species, which are connected through 3680 reactions. The connnectivity of components in complexes is explicit in the graphical representation, unlike for the string representation used in earlier work [16].

Are all these species and reactions important? Recent work indicates that while only a small portion of the FcεRI network is active for a particular set of model parameters (concentrations and rate constants), the active portion depends on the parameter values and activity can be shifted [22]. Reduced models can be found that reproduce predictions of the full model; however, the predictions of these models, relative to the full model, become inaccurate when parameter values are varied over moderate ranges. For details, see [22]. Others are also studying model reduction in the context of signal transduction [23].

3.2 Ligand-Receptor Aggregation with Chains and Rings

Dembo and Goldstein [24] and Posner et al. [25] developed a model for bivalent ligand interaction with a cell-surface bivalent receptor, which is represented in Fig. 6. The ligand is symmetric and its two sites are equivalent. The same holds for the receptor, which is free to diffuse in the two-dimensional membrane surface of a cell. This model was developed to describe ligand-receptor binding and receptor aggregation for the simplest type of antigen capable of aggregating IgE-FcεRI complexes. A complex of IgE and FcεRI can be treated as a bivalent receptor because the complex is long lived and IgE antibody has two antigen-combining sites. This model is more physiological than the simpler binding model considered in the example of the previous

section, which describes bivalent ligand binding to monovalent receptor. The representational conventions proposed here make it easy to combine the two models. The rules of Fig. 6 simply replace reaction rules 1 and 2 in Fig. 5. This simple change results in a combinatorial explosion in the number of possible species and reactions. For example, there are 1854 dimeric receptor states alone.

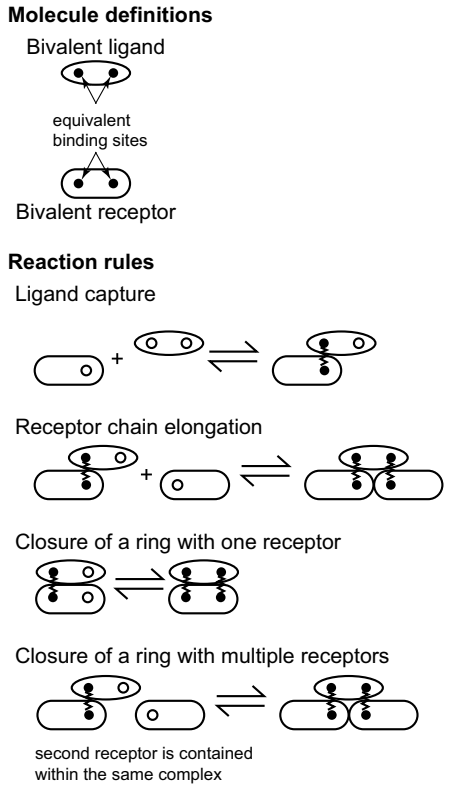

Figure 6: Representation of a model for bivalent ligand interaction with a cell-surface bivalent receptor with chains and rings of receptors included.

3.3 Comparison with a Diagrammatic Representation

Figure 7 shows two representations of a model for the phosphorylation of the retinoblastoma protein (Rb) by a cyclin-dependent kinase [13]. The first representation is diagrammatic and drawn according to the scheme proposed by Kohn [10]. Below, the equivalent rule-based representation is given. Both representations have an unambiguous mathematical interpretation, and both suffer from some of the same drawbacks. For example, both require some investment of time to master, and both are mechanistic rather than functional, making it difficult to identify interactions as stimulatory or inhibitory. A disadvantage of the diagrammatic approach is the need to represent each complex as a separate numbered dot, which is problematic when the number of complexes is large. In contrast, in the rule-based approach, interactions are specified in the form of rules and the complexes implied by these rules can then be identified in an automatic procedure [16, 17]. The rule-based representation is perhaps easier to follow (at least for those already familiar with chemical kinetics), but has the drawback that multiple interactions involving the same component are contained in separate rules[1]. From the point of view of someone trying to develop a mathematical model of the system, the rule-based representation is probably more natural and permits more flexibility and precision in the specification of the reaction network. For example, it is possible to use reaction rules to specify how the molecular context of a component affects the rate at which a reaction takes place, whereas context is difficult to represent in the diagrammatic scheme. We have added boldfaced numbers to the map in Fig. 7 to illustrate how the two representations might be combined, such that reaction rules are used to annotate the arrows of a diagrammatic interaction map. This type of annotation might help to improve the clarity of both maps and rules and resolve ambiguities that often arise in maps by attaching a precise mathematical description to their arrows.

4. CONCLUSIONS

A rule-based representation scheme is comprehensible and precise in certain senses. It is comprehensible in that the information needed to specify or interpret a model is the type of information often available about a system, knowledge of modular protein interactions. In our experience, there is usually a close correspondence between the protein interactions in a system and the reaction rules needed or used to model the system. Importantly, the number of reaction rules needed to represent a system is related to the number of components in the system, which in general is far less than the number of possible chemical species and reactions. Comprehensibility, of course, depends on the ability to read a list of reaction rules. A combination of rules and diagrammatic interaction maps is probably more readable than either type of representation alone. The method of representation is precise in that all the chemical species and reactions implied by specified protein interactions, in the form of rules, are considered. However, when a reaction rule is introduced, a class of reactions is defined, and within this class, the rate of a reaction is sensitive to only specified aspects of molecular context and there is a risk that critical details might be overlooked. Nevertheless, this simplication seems like a good starting point for an iterative cycle of model testing and refinement when one desires to incorporate detail at the level of protein sites and domains.

We were inspired to use graphs and graph rewriting rules to represent signal-transduction systems by the use of graphs and graph rewriting rules to model other types of systems [26, 27, 28]. The advance allowed by the conventions introduced here, relative to earlier rule-based representation [16], is the ability to track the connectivity of components in complexes systematically and explicitly. This ability is important in part because of the complicated polymer-like aggregates that can form through interactions among proteins that contain multiple sites of interaction (Fig. 6) [5]. It is also important if one wishes to adjust the rates of signaling reactions based on the stereochemical properties of reactants. For example, one might wish to make the rate of a reaction depend on the distance between an enzyme and a substrate within a complex, where distance might be

[1]Kitano [12] has proposed a fix to this problem: process diagrams, which each represent a sequence of reaction events. However, multiple process diagrams are needed to account for all possible routes through a branched reaction network.

measured by the number of edges connecting the enzyme and substrate. The cost of explicitly tracking the connectivity of components is the need to find subgraph isomorphisms in graph rewriting steps. Straightforward algorithms exist for finding subgraph isomorphisms [18], but they can be computationally expensive. Fortunately, we expect that most problems will involve small graphs, for which standard methods are effective and feasible.

The conventions introduced here might be extended in several ways. For example, we considered only bonds between components that are affected by signaling (i.e., bonds that can be formed or broken through the application of a reaction rule) and with one exception (Fig. 7), only binary interactions between components. Later, it may be convenient to introduce edge labels to distinguish, for example, between edges that can and cannot be added or removed through graph rewriting. This might facilitate representation of the internal connectivity of the components of a molecule. It may also be convenient to introduce the concept of valence to facilitate the representation of ternary or higher-order interactions between components.

We have presented representational tools that, in principle, can be used to develop an initial mathematical model for any network of proteins for which knowledge of protein-protein interactions is available. This type of knowledge is now being rapidly generated and catalogued in electronic databases. We believe mathematical modeling, and methods of representation like the one presented here, will play an important role in determining how these interactions affect the behavior of a cell. We note that the development of tools for representing and modeling complex biological systems is an active area of research and much work has been done that is related to the work reported here [29, 30, 31, 32].

5. ACKNOWLEDGEMENTS

This work was supported by grants GM35556 and RR18754 from the National Institutes of Health and by the Department of Energy through contract W-7405-ENG-36. We thank Andrew Finney and Paul Loriaux for helpful discussions.

6. REFERENCES

[1] T. Hunter. Signaling—2000 and beyond. *Cell* 100:113–127, 2000.

[2] T. Pawson and P. Nash. Assembly of cell regulatory systems through protein interaction domains. *Science* 300:445–452, 2003.

[3] M. Ptashne and A. Gann. *Genes & Signals.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 2002.

[4] D. Bray. Molecular prodigality. *Science* 299:1189–1190, 2003.

[5] W. S. Hlavacek, J. R. Faeder, M. L. Blinov, A. S. Perelson, and B. Goldstein. The complexity of complexes in signal transduction. *Biotechnol. Bioeng.* 84:783–794, 2003.

[6] J. J. Tyson, K. C. Chen, and B. Novak. Sniffers, buzzers, toggles and blinkers: dynamics of regulatory and signaling pathways in the cell. *Curr. Opin. Cell Biol.* 15:221–231, 2003.

[7] H. S. Wiley, S. Y. Shvartsman, and D. A. Lauffenburger. Computational modeling of the EGF-receptor system: a paradigm for systems biology. *Trends Cell Biol.* 13:43–50, 2003.

[8] B. Goldstein, J. R. Faeder, and W. S. Hlavacek. Mathematical and computational models of immune-receptor signalling. *Nat. Rev. Immunol.* 4:445–456, 2004.

[9] M. Hucka, A. Finney, H. M. Sauro, H. Bolouri, J. C. Doyle, H. Kitano, et al. The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics* 19:524–531, 2003.

[10] K. W. Kohn. Molecular interaction maps as information organizers and simulation guides. *Chaos* 11:84–97, 2001.

[11] R. Maimon and S. Browning. Diagrammatic notation and computational structure of gene networks. In *Proceedings of the Second International Conference on Systems Biology* (T.-M. Yi, M. Hucka, M. Morohashi, and H. Kitano, eds.), pages 311–317, Omnipress, Madison, WI, 2001.

[12] Kitano, H. A graphical notation for biochemical networks. *Biosilico* 1(5):169–176, November 2003.

[13] M. I. Aladjem, S. Pasa, S. Parodi, J. N. Weinstein, Y. Pommier, and K. W. Kohn. Molecular interaction maps—diagrammatic graphical language for bioregulatory networks. *Sci. STKE* 2004(222):pe8, March 2004.

[14] B. Goldstein, J. R. Faeder, W. S. Hlavacek, M. L. Blinov, A. Redondo, and C. Wofsy. Modeling the early signaling events mediated by FcεRI. *Mol. Immunol.* 38:1213–1219, 2002.

[15] J. R. Faeder, W. S. Hlavacek, I. Reischl, M. L. Blinov, H. Metzger, A. Redondo, C. Wofsy, and B. Goldstein. Investigation of early events in FcεRI-mediated signaling using a detailed mathematical model. *J. Immunol.* 170:3769–3781, 2003.

[16] M. L. Blinov, J. R. Faeder, B. Goldstein, and W. S. Hlavacek. BioNetGen: software for rule-based modeling of signal transduction based on the interactions of molecular domains. *Bioinformatics* (in press) doi:10.1093/bioinformatics/bth378, June 2004.

[17] J. R. Faeder, M. L. Blinov, B. Goldstein, and W. S. Hlavacek. Rule-based modeling of biochemical networks. *Complexity* (in press).

[18] J. R. Ullmann. An algorithm for subgraph isomorphism. *J. ACM* 23:31–42, 1976.

[19] H. Dörr. *Efficient Graph Rewriting and Its Implementation*, v. 922 of Lecture Notes in Computer Science. Springer-Verlag, New York, 1995.

[20] B. D. McKay. Practical graph isomorphism. *Congressus Numerantium* 30:45–87, 1981. Also see `http://cs.anu.edu.au/~bdm/nauty/`

[21] J.-L. Faulon and A. G. Sault. Stochastic generator of chemical structure. 3. Reaction network generation. *J. Chem. Inf. Comput. Sci.* 41:894–908, 2001.

[22] J. R. Faeder, M. L. Blinov, B. Goldstein, and W. S. Hlavacek. Combinatorial complexity and dynamical restriction of network flows in signal transduction. `arXiv:q-bio.MN/0411021`, submitted to *Syst. Biol.*

[23] H. Conzelmann, J. Saez-Rodriguez, T. Sauter, E. Bullinger, F. Allgöwer, and E. D. Gilles. Reduction of

mathematical models of signal transduction networks: simulation-based approach applied to EGF receptor signalling. *Syst. Biol.* 1:159–169.

[24] M. Dembo and B. Goldstein. Theory of equilibrium binding of symmetric bivalent haptens to cell surface antibody: application to histamine release from basophils. *J. Immunol.* 121:345–353, 1978.

[25] R. G. Posner, C. Wofsy, and B. Goldstein. The kinetics of the bivalent ligand-bivalent receptor aggregation: ring formation and the breakdown of the equivalent site approximation. *Math. Biosci.* 126:171–190, 1995.

[26] J.-L. Faulon. Isomorphism, automorphism partitioning, and canonical labeling can be solved in polynomial time for molecular graphs. *J. Chem. Inf. Comput. Sci.* 38:432–444, 1998.

[27] G. Benkö, C. Flamm, and P. F. Stadler. A graph-based toy model of chemistry. *J. Chem. Inf. Comput. Sci.* 43:1085–1093, 2003.

[28] E. Klavins, R. Ghrist, and D. Lipsky. Graph grammars for self-assembling robotic systems. In *Proceedings of the 2004 IEEE International Conference on Robotics and Automation* 5:5293–5300, 2004.

[29] A. Regev and E. Shapiro. The π-calculus as an abstraction for biomolecular systems. *Nat. Comp. Ser.* pp. 219–266, 2004.

[30] V. Danos and C. Laneve. Formal molecular biology. *Theor. Comput. Sci.* 325:69–110, 2004.

[31] C. Talcott, S. Eker, M. Knapp, P. Lincoln, and K. Laderoute. Pathway logic modeling of protein functional domains in signal transduction. *Pac. Symp. Biocomput.* pp. 568–580, 2004.

[32] B. R. Franza. From play to laws: language in biology. *Sci. STKE* pe9, 2004.

James R. Faeder received his Ph.D. in chemical physics from the University of Colorado and did postdoctoral work at Colorado State University, the Weizmann Institute of Science, and Los Alamos National Laboratory, where he is now a Technical Staff Member. His research has involved developing and analyzing detailed models of complex chemical systems. He is now focused on modeling biochemical signaling networks.

Michael L. Blinov received his Ph.D. in mathematics from the Weizmann Institute of Science. During his graduate work he interned in the Theoretical Biology and Biophysics Group at Los Alamos National Laboratory, where he is now a Postdoctoral Research Associate.

William S. Hlavacek received his Ph.D. in chemical engineering at the University of Michigan and did postdoctoral work at Los Alamos National Laboratory, where he is now a Technical Staff Member. He is also affiliated with the Department of Biology and the Center for Evolutionary and Theoretical Immunology at the University of New Mexico. He has worked on modeling bacterial gene regulation, multivalent ligand-receptor binding, HIV dynamics, and eukaryotic signal transduction.

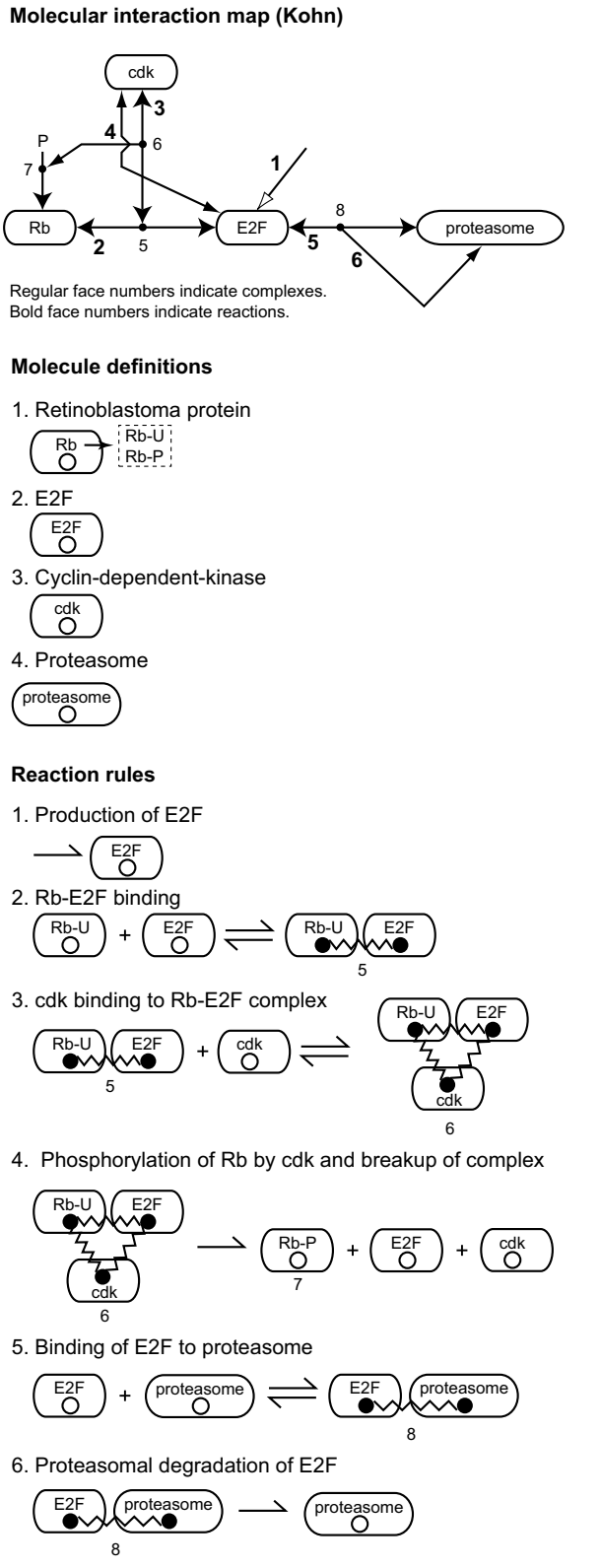

Figure 7: A formal diagrammatic map and the corresponding set of graphical reaction rules.